

IMPLEMENTASI NAIVE BAYES DALAM MEMPREDIKSI PENYAKIT *DIABETES MELLITUS*

Qois Al'Ariq¹⁾; Ginanjar Setyo Permadi²⁾; Chamdan Mashuri³⁾; Anita Andriani⁴⁾

^{1,2,3,4)} Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari Tebuireng Jombang

Alamat e-mail^{1,2,3,4)}

goisalarig@mhs.unhasy.ac.id¹, ginanjarsetyo@unhasy.ac.id²,
chamdanmashuri@unhasy.ac.id³, anitaandriani@unhasy.ac.id⁴

Abstrak

Diabetes melitus adalah salah satu penyakit yang secara signifikan mempengaruhi populasi global. Penelitian ini mengeksplorasi penerapan Naive Bayes untuk memprediksi diabetes melitus. Permasalahan yang dihadapi antara lain kompleksitas data klinis, keragaman fitur, dan kebutuhan akan prediksi yang akurat. Solusi yang diusulkan adalah dengan menggunakan algoritma klasifikasi Naive Bayes yang memanfaatkan asumsi sederhana namun kuat tentang independensi fitur. Sistem ini dijelaskan dengan langkah-langkah yang melibatkan pra-pemrosesan data, partisi kumpulan data, pelatihan model Naive Bayes, dan evaluasi kinerja. Dataset yang digunakan meliputi usia, jenis kelamin, berat badan, HbA1C, gula darah puasa. Hasil dan pengujian menunjukkan bahwa model Naive Bayes dapat memberikan prediksi diabetes melitus yang cukup akurat, dengan kinerja yang dinilai melalui metrik evaluasi seperti recall, akurasi, skor F1, dan presisi. Kesimpulannya, penerapan Naive Bayes merupakan pendekatan yang cukup efektif dalam memprediksi diabetes melitus. Hasil Kinerja metode Naive Bayes pada penelitian ini dinilai cukup akurat, hal ini dibuktikan dalam skor akurasi tertinggi yaitu **85.71%**. Meskipun sederhana, algoritma ini dapat menangani kompleksitas dataset dan memberikan prediksi yang andal.

Kata Kunci: *Diabetes Mellitus, Data Mining, Machine Learning, Klasifikasi, dan Naive Bayes.*

Abstract

*Diabetes mellitus is a illness that significantly affects the global population. This study explores the implementation of Naive Bayes to predict diabetes mellitus. The problems faced include the complexity of clinical datasets, feature diversity, and the need for accurate predictions. The proposed solution is to use the Naive Bayes classification algorithm that utilizes a simple but strong assumption about feature independence. The system is described with steps involving data pre-processing, dataset partitioning, Naive Bayes model training, and performance evaluation. The datasets used include age, gender, weight, HbA1C, fasting blood sugar. The results and testing show that the Naive Bayes model can provide fairly accurate predictions for diabetes mellitus, with performance assessed through evaluation metrics such as , recall, accuracy, F1-score, and precision. In conclusion, the implementation of Naive Bayes is a fairly effective approach to predicting diabetes mellitus. In this study, The performance of the Naive Bayes method is considered quite accurate, this is evidenced in the highest accuracy score of **85.71%**. Despite its simplicity, this algorithm can handle the complexity of the dataset and provide reliable predictions.*

Keywords: *Diabetes Mellitus, Data Mining, Machine Learning, Classification, and Naive Bayes.*

PENDAHULUAN

Diabetes atau *Diabetes Mellitus* adalah suatu gangguan pada kondisi kesehatan yang ditandai dengan adanya kadar gula darah yang melebihi batas nilai normal yaitu lebih dari atau sama dengan 200 mg/dl dan juga kadar gula darah puasa yang bernilai sama dengan atau lebih dari 126 mg/dl. Penyakit *Diabetes* ini adalah suatu penyakit yang mematikan, hal ini disebabkan karena penderitanya telat mengetahui gejala tersebut dan berakibat komplikasi, karena itulah *diabetes* ini dikenal juga sebagai *silent killer* (Febriyanti, 2021). Komplikasi pada lebih dari satu bagian tubuh dan juga bisa menambah risiko kematian dini adalah akibat fatal dari penyakit ini. Komplikasi yang mungkin terjadi pada penderita diabetes antara lain kehilangan penglihatan, gagal ginjal, kerusakan saraf, dan amputasi kaki. Peningkatan risiko serangan jantung dan stroke dapat meningkat dua hingga tiga kali lipat apabila diderita oleh orang dewasa. Peningkatan risiko kematian janin dan komplikasi lainnya pada masa kehamilan disebabkan karena diabetes tidak dapat terkontrol dengan baik. (WHO, 2020).

Total pengidap penyakit *diabetes* pada tahun 2019 di dunia adalah sebanyak 463 juta penderita dan di tahun 2045 diprediksi akan bertambah naik dengan total penderita meningkat sebesar 51% atau sebanyak 700 juta. (Webber, 2013). Pada data serupa, Indonesia menempati urutan ke tujuh negara di dunia dengan pengidap penyakit *diabetes* terbanyak dari total 211 negara yang tercatat dalam IDF dan total penderita sejumlah 10,7 juta jiwa. Selain itu, untuk wilayah Pasifik Barat, negara Indonesia menempati posisi ke dua di bawah negara China dari total 36 negara yang tercatat

sebanyak 6,2% dari total jumlah orang dewasa yaitu sebesar 172,2 juta orang pada tahun 2019 (Webber, 2013). Mayoritas penderita *diabetes* bertempat tinggal di negara berpendapatan rendah dan berpendapatan menengah dan juga setiap tahunnya, kematian disebabkan oleh *diabetes* sebanyak 1,6 juta. Hal tersebut menjadikan penyakit ini sebagai salah satu penyebab utama kematian dari sepuluh penyakit di dunia. (IDF, 2020). Indonesia menempati posisi ke-7 di dunia pada tahun 2019, setelah Amerika Serikat, China, Pakistan, India, Meksiko, dan Brazil, sebagai negara dengan total penderita *diabetes* tertinggi, dengan jumlah penderita sebanyak 10,7 juta orang (Saeedi dkk., 2019). Angka ini meningkat menjadi 10,8 juta pada tahun 2020 dengan persentase jumlah pasien pengidap penyakit *diabetes* sebesar 6,2% dan pada tahun 2045 jumlah pengidap penyakit ini di negara Indonesia diperkirakan akan meningkat menjadi 16,7 juta. (Risksdas, 2018).

Suatu metode yang dipergunakan untuk mendeteksi penyakit *diabetes* yaitu dengan cara menggunakan algoritma *Machine Learning* (Jashwanth Reddy dkk., 2020). Dalam berbagai bidang, khususnya bidang kesehatan, metode ini telah banyak digunakan. (Krishnamoorthi, 2023). Metode *Naive Bayes* adalah suatu algoritma dalam pengelompokan atau klasifikasi pada *Machine Learning*. Algoritma ini dapat memprediksi suatu keanggotaan kelas yang terdapat di suatu kelas dengan memanfaatkan teori probabilitas. *Teorama Bayes* merupakan sebuah dasar dari algoritma *Naive Bayes* yang digunakan pertama kali oleh *Thomas Bayes* pada abad ke-18 dalam teori probabilitas dan keputusan. (Fitriyani, 2021). Metode *Naive Bayes* termasuk juga dalam teknik dasar pembelajaran mandiri dalam artian teknik dasar pembelajaran tersebut memiliki cara pembelajaran yang berbeda sehingga penerapan metode *naive bayes* didasarkan pada kebutuhan dari *Machine Learning* itu sendiri memiliki jenis pendekatan berupa *Unsupervised Learning* dan *Supervised Learning*. (Santoso dkk., 2021)(Nengsih, 2019). Pada penelitian ini, peneliti akan mempergunakan jenis *Supervised Learning*, karena pendekatan tersebut mempergunakan data yang telah memiliki label untuk memprediksi hasil output sesuai input yang telah diberikan (Haris dkk., 2022). Contoh dari pengaplikasian pendekatan *Supervised Learning* yaitu sistem rekomendasi yang memprediksi serial TV atau film yang mungkin diminati oleh pengguna berdasarkan pada data yang telah diperoleh mengenai film yang telah dilihat. (Savitri & Nursalim, 2023). Dengan adanya sistem prediksi dengan menggunakan pendekatan *Supervised Learning* tersebut, sebelum dilakukan pemeriksaan medis tahap lanjut, pengecekan *diabetes* dapat dilakukan lebih dini oleh seseorang dengan menggunakan bantuan *Machine Learning*, apabila prediksi memperlihatkan bahwa pengguna tersebut terdiagnosa penyakit *diabetes*, maka akan dilakukan pemeriksaan medis lebih lanjut. Pengecekan dengan metode tersebut bagi rakyat sangat terjangkau karena tidak membutuhkan biaya dan dapat diakses dari mana saja.

METODE

Diabetes melitus adalah suatu gangguan metabolisme kronis atau gangguan autoimun kronis yang diakibatkan oleh cara kerja atau sekresi insulin, atau kedua-duanya yang tidak normal dan memiliki ciri kadar glukosa dalam darah terlihat cukup tinggi. (Susilawati, 2018). *Diabetes* melitus tipe 2 terjadi pada saat tubuh manusia tidak lagi dapat untuk membuat insulin dengan cukup sehingga glukosa dari darah tidak mampu diangkat ke sel-sel tubuh (F. dan B. Aris, 2019). Gangguan proses transportasi glukosa disebabkan oleh ketidakmampuan sel dalam merespon insulin sehingga terjadi kadar glukosa dalam darah menjadi tinggi (Kemenkes RI, 2018).

Data mining merupakan suatu definisi yang dipakai sebagai sarana dalam menganalisis data di dalam *database* (Rizki dkk., 2022). Tahapan yang menerapkan *Machine Learning*, teknik statistik, *artificial intelligent*, dan matematika yang digunakan sebagai sarana dalam mengolah informasi yang berguna dan informasi yang terkait dari *database* yang besar disebut *Data mining* (Anggada Maulana, 2018). Tahapan pencarian menggunakan *pattern* yang tersembunyi atau (*hidden patern*) yang berupa pengetahuan atau (*knowledge*) yang sebelumnya tidak diketahui, dari sebuah data yang berada di dalam data *warehouse*, *database*, atau media untuk menyimpan informasi yang berbeda merupakan deskripsi umum dari *data mining*. (Wahono, 2023).

Pembelajaran mesin atau *Machine Learning* adalah penerapan *Artificial Intelligent* yang memberikan sistem kemampuan belajar secara otomatis dari sebuah data untuk melakukan pekerjaan tertentu tanpa harus diprogram secara eksplisit (Roihan dkk., 2020). *Machine Learning* juga adalah bidang ilmu komputer yang mempelajari teknik dan algoritma yang menghasilkan sebuah penyelesaian secara otomatis pada masalah yang kompleks dan susah dilakukan dengan menerapkan metode pemrograman konvensional (Hasydna & Dinata, 2020).

Klasifikasi merupakan sebuah prosedur yang digunakan untuk menemukan sebuah fungsi atau model untuk membedakan dan menjelaskan konsep yang memiliki tujuan guna dapat memperkirakan atau memprediksi suatu kelas yang berasal dari suatu objek dimana kelas belum pernah diketahui sebelumnya. (Hendrian, 2018).

Naive Bayes adalah suatu metode klasifikasi probabilistik yang sederhana dalam menghitung beberapa kemungkinan dengan cara menambahkan kombinasi serta frekuensi nilai yang didasarkan pada dataset (Maulidah dkk., 2021).

Mempergunakan Algoritma Bayes dan berasumsi seluruh komponen berdiri sendiri atau tidak ketergantungan yang diberikan pada suatu nilai variabel kelas. Deskripsi yang berbeda tentang *Naive Bayes* yaitu proses klasifikasi dengan cara statistik dan probabilitas, dicetuskan oleh ilmuwan yang berasal dari Inggris bernama Thomas Bayes, yakni dapat prediksi masa depan menggunakan peluang berdasarkan pengalaman dari masa sebelumnya (Rachman & Handayani, 2021). Berdasarkan definisi penyederhanaan, apabila diberi suatu nilai *output*, maka nilai atribut saling tidak tetikat satu sama lain. (Widodo dkk., 2021). Persamaan dari *Teorema Bayes* adalah :

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Confusion Matrix merupakan metode yang digunakan dalam penilaian kinerja model klasifikasi pada data mining dengan memberikan informasi tentang kebenaran hasil penilaian prediksi. *Matrix* terdiri dari *True Positive (TP)*, *False Positive (FP)*, *True Negative (TN)*, dan *False Negative (FN)* digunakan sebagai perhitungan *Accuration, Precision, Recall* dan *F1 Score*. (Mutiarasari, 2019)

$$ACCURATION = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$PRECISION = \frac{TP}{TP + FP} \quad (3)$$

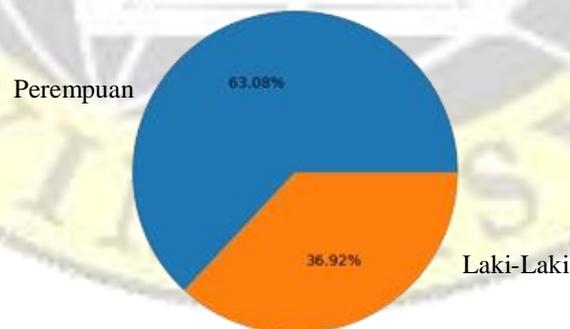
$$RECALL = \frac{TP}{TP + FN} \quad (4)$$

$$F1\ SCORE = 2X \frac{PRECISION \times RECALL}{PRECISION + RECALL} \quad (5)$$

HASIL DAN PEMBAHASAN

A. Data Pengujian

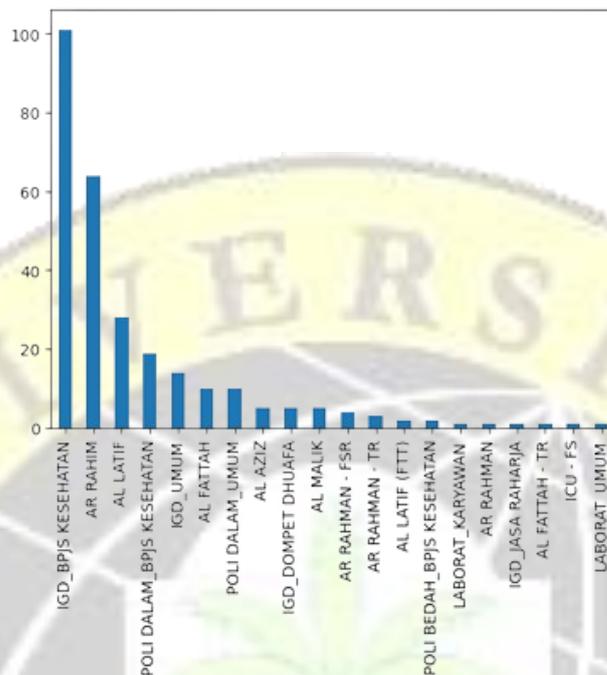
Data yang diproses oleh peneliti yaitu berupa data pasien Rumah Sakit Hasyim Asy'ari Jombang berupa Tanggal Pemeriksaan, Hasil HbA1C, Satuan HbA1C, Nilai Rujukan HbA1C, Hasil HbA1C > Nilai Rujukan, Hasil Puasa, Satuan Glukosa Puasa, Nilai Rujukan, Puasa > Nilai Rujukan, Hasil Terdiagnosa Diabetes, Nama Ruangan, Umur, Jenis Kelamin, dan Berat Badan. Data pasien Rumah Sakit Hasyim Asy'ari Jombang yang didapatkan saat pengumpulan data berjumlah 279 pasien, dengan jumlah pasien perempuan 176 orang dan pasien laki-laki 103 orang.



Gambar 1. Diagram Jumlah Pasien Rumah Sakit Hasyim Asy'ari

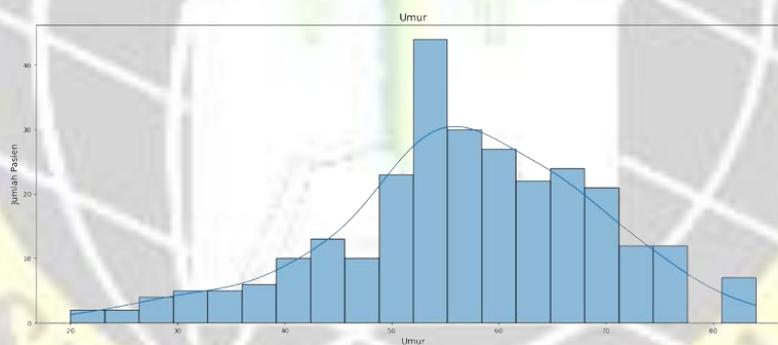
Masing-masing pasien berasal dari ruangan yang berbeda-beda di Rumah Sakit Hasyim Asy'ari Jombang. Dengan rincian sebagai berikut : Igd_Bpjs Kesehatan (101), Ar Rahim(64), Al Latif(28), Poli Dalam_Bpjs Kesehatan (19), Igd_Umum (14), Al Fattah (10), Poli Dalam_Umum (10), Al Aziz (5), Igd_Dompot Dhuafa (5), Al Malik (5), Ar

Rahman – Fsr (4), Ar Rahman – Tr (3), Al Latif (Ftt) (2), Poli Bedah_Bpjs Kesehatan (2), Laborat_Karyawan (1), Ar Rahman(1), Igd_Jasa Raharja (1), Al Fattah - Tr (1), Icu – Fs (1), Laborat_Umum (1). Jumlah pasien per ruangan yang bisa dilihat pada Gambar 2 di bawah ini :



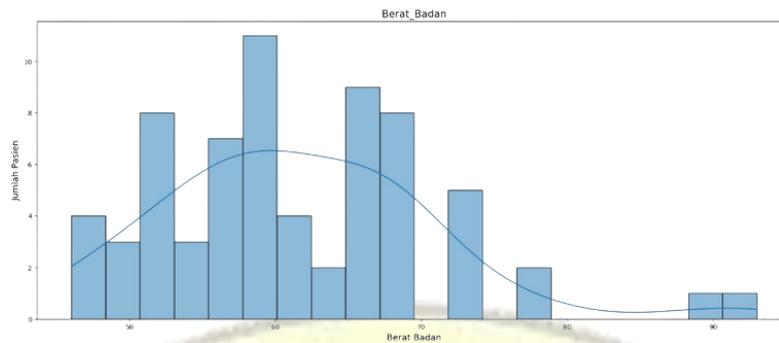
Gambar 2. Nama Ruangan

Pasien juga berasal dari latar belakang umur yang berbeda, seperti dalam Gambar 3. Sebagian besar pasien berada pada rentang usia 50-60 tahun.



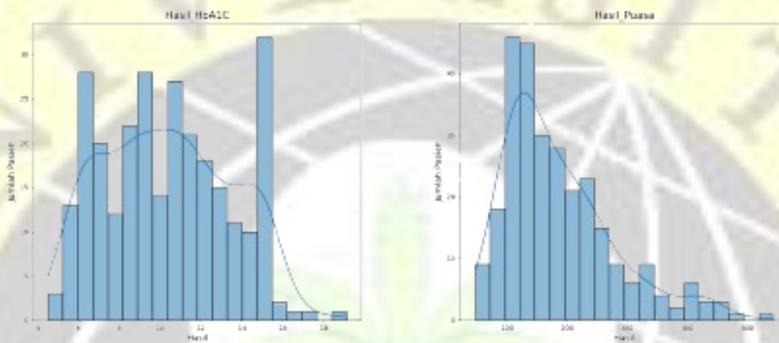
Gambar 3. Usia Pasien

Berat badan pasien pun berbeda-beda, seperti dalam Gambar 4. Sebagian besar pasien berada pada rentang berat badan 50-70 kg.



Gambar 4. Berat Badan Pasien

Pada Gambar 5, dapat dilihat persebaran data Hasil HbA1C dan data Hasil Glukosa Puasa dalam bentuk *bar plot*.



Gambar 5. Hasil Pemeriksaan

B. Preprocessing

Pada tahapan ini dijalankan tahap perubahan data yang masih mentah menjadi data yang sudah siap diolah dan akan mempermudah dalam proses analisa data.

1. Import Library

Library adalah sebuah modul atau paket yang diinstal. Import library memiliki tujuan untuk dapat memproses fungsi-fungsi yang dibutuhkan dalam pembangunan model prediksi. Beberapa library yang diperlukan dalam penelitian ini berfungsi untuk proses komputasi numerik (NumPy dan Math), perhitungan yang lebih kompleks (Scipy), mengacak dan memilih acak elemen list (Random), menampilkan jenis visualisasi (Matplotlib dan Seaborn), pemrosesan data (Pandas), dan pemodelan Machine Learning (Scikit-Learn).

2. Menampilkan Dataset

Tahapan selanjutnya sebelum mengolah data yaitu menampilkan dataset yang telah didapatkan dari proses observasi di lapangan yang telah disimpan dalam ekstensi .xlsx

3. Null Value Dataset

Dilakukan pengecekan *null value* karena di dalam dataset berpeluang terdapat data yang belum terisi.

4. Menghapus Kolom

Proses selanjutnya yakni menghilangkan beberapa kolom yang terdapat pada dataset dikarenakan kolom tersebut telah tidak diperlukan lagi dalam pemrosesan data yang nantinya akan digunakan untuk memprediksi penyakit *diabetes mellitus*

5. Scalling Data

Metode *scalling data* yang dipakai pada penelitian ini yaitu Standard Scaler. *Scalling data* digunakan untuk menormalkan rentang variabel independen atau fitur data.

C. Naïve Bayes

1. Memisahkan data menjadi data *test* dan data *train*
2. Mengklasifikasikan data menggunakan metode *Naïve Bayes*
3. Menyesuaikan data ke dalam model
4. Memprediksi data menggunakan metode *Naïve Bayes*
5. Menampilkan hasil klasifikasi
6. Menguji model

Menguji model dalam memprediksi penyakit *diabetes mellitus* menggunakan metode *Naïve Bayes*. Menginputkan HbA1C, gula darah puasa, usia, jenis kelamin, dan berat badan. Apabila hasil *output* menunjukkan

0, maka diprediksi tidak terkena diabetes dan sebaliknya apabila hasil *output* menunjukkan 1, maka diprediksi terkena penyakit *diabetes mellitus*.

```
print(bayes_model.predict([[10.8,101,59,1,50]]))
```

[36]

... [0]

Gambar 6. Hasil Uji Model (1)

```
print(bayes_model.predict([[6.8,228,73,0,60]]))
```

[37]

... [1]

Gambar 7. Hasil Uji Model (2)

D. Menampilkan hasil prediksi melalui *website*

1. *Import Library*

Library yang dibutuhkan untuk menampilkan hasil prediksi via *website* yaitu *pickle* dan *streamlit*. *Pickle* digunakan untuk menyimpan dan memanggil modul. Kemudian *library streamlit* digunakan untuk membuat *interface* dengan menggunakan bahasa pemrograman Python.

2. Memanggil File *Pickle*

File *pickle* yang di dalamnya sudah tersimpan metode *Naïve Bayes* dipanggil agar dapat memprediksi penyakit *Diabetes Mellitus* melalui *website*.

3. Melakukan Prediksi

Dilakukan prediksi dengan metode *Naïve Bayes* yang sebelumnya sudah disimpan melalui *pickle*.

4. Tampilan *Website*

Website prediksi penyakit *diabetes mellitus* ini terdapat beberapa *page*, antara lain : *Page home*, *data frame demo*, *plotting demo*, dan *page* untuk prediksi diabetes. Tampilan *website* pada beberapa *page* bisa dilihat di gambar-gambar di bawah ini :

1. *Page Home*

Yang pertama yaitu halaman awal *website* prediksi *diabetes mellitus* ini ditampilkan penjelasan mengenai penyakit *diabetes mellitus* berupa penyebab penyakit diabetes, tipe diabetes, faktor yang menyebabkan beresiko terkena penyakit diabetes, dan grafik peningkatan kasus diabetes di Indonesia setiap tahun. Berikut ini ditampilkan halaman *home* dari *website* seperti gambar-gambar di bawah ini :



Gambar 8. *Page Home* (1)



Gambar 9. Page Home (2)



Gambar 10. Page Home (3)

2. Page Data Frame Demo

Pada halaman *website* ini ditampilkan dataset yang diperoleh dari data pasien Rumah Sakit Hasyim Asy'ari. Data tersebut berupa Tanggal Pemeriksaan, Hasil HbA1C, Satuan HbA1C, Nilai Rujukan HbA1C, Hasil HbA1C > Nilai Rujukan, Hasil Gula Darah Puasa, Satuan Glukosa Puasa, dan Nilai Rujukan Glukosa Puasa.

No	Tanggal Pemeriksaan	Hasil HbA1C	Satuan HbA1C	Nilai Rujukan HbA1C	Hasil HbA1C > Nilai Rujukan	Hasil Puasa	Satuan Glukosa Puasa	Nilai Rujukan
1	30/04/2024	10,6	%	5,7	1	453	mg/dl	70 - 105
2	30/04/2024	12,5	%	5,7	1	439	mg / dl	70 - 105
3	30/04/2024	6,5	%	5,7	1	111	mg / dl	70 - 105
4	27/04/2024	10,3	%	5,7	1	354	mg / dl	70 - 105
5	27/04/2024	10,5	%	5,7	1	135	mg / dl	70 - 105
6	25/04/2024	11,3	%	5,7	1	120	mg/dl	70 - 105
7	24/04/2024	8,6	%	5,7	1	123	mg / dl	70 - 105
8	23/04/2024	10	%	5,7	1	113	mg / dl	70 - 105
9	23/04/2024	10,6	%	5,7	1	202	mg / dl	70 - 105
10	23/04/2024	6,3	%	5,7	1	86	mg/dl	70 - 105
11	23/04/2024	7,1	%	5,7	1	184	mg / dl	70 - 105

Gambar 11. Page Data Frame Demo

3. Page Plotting Demo

Pada halaman *website* ini ditampilkan beberapa plot atau grafik dari data pasien di Rumah Sakit Hasyim Asy'ari. Grafik yang ditampilkan pada gambar-gambar di bawah ini berupa visualisasi data dari Nama Ruang, Jenis Kelamin, Umur, dan Berat Badan.



Gambar 12. Page Plotting Demo (1)



Gambar 13. Page Plotting Demo (2)

4. Page Prediksi

Pada halaman *website* ini ditampilkan prediksi penyakit *diabetes mellitus* yang sebelumnya telah dibuat dengan menggunakan metode Naïve Bayes. Pada halaman ini pengguna diminta untuk menginputkan usia, berat badan, HbA1C, dan Gula Darah Puasa. Pada gambar-gambar di bawah berupa hasil prediksi dari input yang telah dimasukkan oleh user menampilkan hasil terdiagnosa diabetes ataupun tidak terdiagnosa diabetes.

The screenshot shows the 'Prediksi' page of the web application. It contains several input fields: 'Input Usia' (23), 'Input Berat Badan' (100), 'Input HbA1C' (10), and 'Input Gula Darah Puasa' (130). There is a dropdown menu for 'Pilih Jenis Kelamin' set to 'Wanita'. A 'Test Prediksi Diabetes' button is located below the inputs. Below the button, a message box displays 'Anda terprediksi terkena diabetes'. A green 'Success' notification is visible in the bottom right corner.

Gambar 14. Page Prediksi (1)

This screenshot shows the 'Prediksi' page with different input values. The 'Input Usia' is 60, 'Input Berat Badan' is 62, 'Input HbA1C' is 13.2, and 'Input Gula Darah Puasa' is 107. The 'Pilih Jenis Kelamin' dropdown is set to 'Pria'. The 'Test Prediksi Diabetes' button is present. The message box now displays 'Anda terprediksi tidak terkena diabetes'. A green 'Success' notification is visible in the bottom right corner.

Gambar 15. Page Prediksi (2)

E. Pengukuran Akurasi

Pengukuran akurasi dilakukan dengan cara menghitung skor *Precision*, *Accuracy*, *F1-Score*, dan *Recall* dengan menggunakan rumus yang dilakukan dalam beberapa skenario pembagian data *train* dan data *test*. Perbandingan skor *Precision*, *Accuracy*, *F1-Score*, dan *Recall* pada setiap skenario pembagian data *test* dan data *train* dapat dilihat pada tabel berikut ini :

Tabel 1. Pengukuran Akurasi

	70% Test 30% Train	60% Test 40% Train	55% Test 45% Train	50% Test 50% Train	40% Test 60% Train	30% Test 70% Train	25% Test 75% Train	20% Test 80% Train
<i>Accuracy</i>	72.95%	73.80%	73.37%	74.28%	75%	76.19%	85.71%	83.92%
<i>Precision</i>	76%	78%	77%	79%	81%	85%	91%	89%
<i>Recall</i>	84%	83%	83%	82%	81%	79%	87%	86%
<i>F1-Score</i>	80%	80%	80%	80%	81%	81%	89%	88%

Berdasarkan tabel tersebut dapat dilihat bahwa skor akurasi tertinggi yaitu 85.71% pada data test 25% dan data train 75%, skor akurasi terendah sebesar 72.95% pada data test 70% dan data train 30%. Persentase *precision* terbesar yaitu 91% pada 25% data test dan 75% data train, sedangkan persentase *precision* terendah yaitu pada data test 70% dan data train 30%. Kemudian untuk persentase *recall* terbesar yaitu pada 25% data test dan 75% data train sebesar 87%, nilai *recall* terendah yaitu 79% pada 30% data test dan 70% data train. Selanjutnya yang terakhir yaitu *f1-score* tertinggi yaitu pada 25% data test dan 75% data train sebesar 89% dan nilai *f1-score* terendah dengan 80% pada 30% Train 70% Test, 40% Train 60% Test, 45% Train 55% Test, dan 50% Train 50% Test.

Pada penelitian ini didapatkan akurasi tertinggi menggunakan metode Naive Bayes dalam melakukan prediksi penyakit *diabetes mellitus* yaitu sebesar **85.71%**. Dengan hasil prediksi tersebut, penggunaan metode Naive Bayes pada penelitian ini dinyatakan cukup akurat dalam memprediksi penyakit *diabetes mellitus*.

PENUTUP

Simpulan

Berdasarkan pembahasan dan hasil yang dilakukan dalam penelitian tentang “IMPLEMENTASI NAIVE BAYES DALAM MEMPREDIKSI PENYAKIT *DIABETES MELLITUS* (STUDI KASUS RUMAH SAKIT HASYIM ASY'ARI)”, dapat disimpulkan bahwa :

1. Pembuatan sistem deteksi penyakit *diabetes mellitus* menggunakan metode naive bayes pada dataset yang digunakan dalam penelitian ini terbukti metode naive bayes memiliki akurasi cukup tinggi dan efektif dalam mendeteksi penyakit *diabetes mellitus* dengan jumlah data yang besar.
2. Dalam mengukur kinerja metode naive bayes dilakukan beberapa skenario pembagian data *train* dan data *test* (30% Train 70% Test, 40% Train 60% Test, 45% Train 55% Test, 50% Train 50% Test, 60% Train 40% Test, 70% Train 30% Test, 75% Train 25% Test, dan 80% Train 20% Test). Kinerja metode Naive Bayes pada penelitian ini dinilai cukup akurat, hal ini dibuktikan dalam skor akurasi tertinggi yaitu **85.71%** pada data *test* 25% dan data *train* 75%.

Saran

Berdasarkan hasil yang didapatkan dari penelitian ini, maka dapat ditarik saran yang berguna dalam penelitian dan pengembangan berikutnya, yaitu melakukan percobaan menggunakan metode selain Naive Bayes dalam menentukan prediksi penyakit *diabetes mellitus* untuk dapat mengetahui hasil akurasi prediksi menggunakan metode yang lain.

DAFTAR PUSTAKA

WHO. (2020). Diabetes. Whorld Health Oraganization. <https://www.who.int/news-room/fact-sheets/detail/diabetes>

Wahono, R. S. (2023). Data Mining Data mining. In Mining of Massive Datasets (Vol. 2, Issue January 2013). https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

IDF. (2020). Diabetes facts and figures show the growing global burden for individuals, families, and countries. The IDF Diabetes Atlas (2021) reports that 10.5% of the adult population (20-79 years) has diabetes, with almost half unaware that they are living with the. The International Diabetes Federation. <https://idf.org/about-diabetes/diabetes-facts-figures/>

- Riskesdas, T. (2018). Laporan Nasional Riskesdas 2018 (p. 674). Badan Penelitian dan Pengembangan Kesehatan.
- Jashwanth Reddy, D., Mounika, B., Sindhu, S., Pranayteja Reddy, T., Sagar Reddy, N., Jyothsna Sri, G., Swaraja, K., Meenakshi, K., & Kora, P. (2020). WITHDRAWN: Predictive machine learning model for early detection and analysis of diabetes. *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2020.09.522>
- Krishnamoorthi, R. (2023). A Novel Diabetes Healthcare Disease Prediction Framework Using Machine Learning Techniques. *National Library of Medicine*. <https://doi.org/10.1155/2022/1684017>
- Fitriyani, F. (2021). Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 7(2), 61–69. <https://doi.org/10.25077/teknosi.v7i2.2021.61-69>
- Santoso, P., Abijono, H., & Anggreini, N. L. (2021). ALGORITMA SUPERVISED LEARNING DAN UNSUPERVISED. 4(2), 315–318.
- Nengsih, W. (2019). Analisa Akurasi Permodelan Supervised Dan Unsupervised Learning Menggunakan Data Mining. *Sebatik*, 23(2), 285–291. <https://doi.org/10.46984/sebatik.v23i2.771>
- Savitri, L., & Nursalim, R. (2023). Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning. *Diophantine Journal of Mathematics and Its Applications*, 2(01), 30–36. <https://doi.org/10.33369/diophantine.v2i01.28260>
- Aris, F. dan B. (2019). Penerapan Data Mining untuk Identifikasi Penyakit Diabetes Melitus dengan Menggunakan Metode Klasifikasi. *Router Research*, 1(1), 1–6.
- Kemendes RI. (2018). Hari diabetes. Infodatin Pusat Data Dan Informasi Kementerian Kesehatan RI. <https://pusdatin.kemkes.go.id/article/view/19041500002/hari-diabetes-sedunia-2018.html>
- Rizki, R., Athallah, R., Cholissodin, I., & Adikara, P. P. (2022). Prediksi Potensi Pengidap Penyakit Diabetes berdasarkan Faktor Risiko Menggunakan Algoritme Kernel K-Nearest Neighbor. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 6(8), 3777–3785. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11439>
- Anggada Maulana. (2018). Konsep Dasar Data Mining. *Konsep Data Mining*, 1, 1–16.
- Wahono, R. S. (2023). Data Mining Data mining. In *Mining of Massive Datasets (Vol. 2, Issue January 2013)*. https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- HASYDNA, N., & DINATA, R. K. (2020). *Machine Learning.Pdf*.
- Meilina, P. (2022). Penerapan Data Mining Dengan Metode Kalsifikasi Menggunakan. *Teknologi*, 7(March), 10–20.
- Maulidah, N., Supriyadi, R., Utami, D. Y., Hasan, F. N., Fauzi, A., & Christian, A. (2021). Prediksi Penyakit Diabetes Melitus Menggunakan Metode Support Vector Machine dan Naive Bayes. *Indonesian Journal on Software Engineering (IJSE)*, 7(1), 63–68. <https://doi.org/10.31294/ijse.v7i1.10279>
- Rachman, R., & Handayani, R. N. (2021). Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM. *Jurnal Informatika*, 8(2), 111–122. <https://doi.org/10.31294/ji.v8i2.10494>
- Widodo, Y. B., Anggraeni, S. A., & Sutabri, T. (2021). Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer*, 7(1), 112–123. <https://doi.org/10.37012/jtik.v7i1.507>